UCRL-ID-144292

Description of the ARM Operational Objective Analysis System

M. Zhang, S. Xie, R.T. Cederwall, J.J. Yio

May 1, 2001

U.S. Department of Energy
Lawrence Livermore National Laboratory

DISCLAIMER

This work was performed under the auspices of the U. S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

Description of the ARM Operational Objective Analysis System

May 2001

Minghua Zhang

State University of New York at Stony Brook, Stony Brook, NY, USA

Shaocheng Xie
Richard T. Cederwall
J. John Yio

Lawrence Livermore National Laboratory, Livermore, CA, USA

Work supported by the U.S. Department of Energy,
Office of Energy Research, Office of Health and Environmental Research

Contents

Figures

Tables

1. Introduction

This report describes the ARM (Atmospheric Radiation Measurement) operational variational objective analysis system. It is currently used to process the data collected from the ARM Intensive Operational Periods (IOPs) for driving and evaluating physical parameterizations in climate models. The analysis system was originally developed by Zhang and Lin (1997) at State University of New York at Stony Brook and was migrated to the Lawrence Livermore National Laboratory (LLNL) as the ARM operational objective analysis system in May 1999. In contrast with previous objective analysis (e.g., Barnes, 1964; O'Brien, 1970; Lin and Johnson, 1996), the ARM objective analysis used the constrained variational analysis method developed by Zhang and Lin (1997), in which the atmospheric state variables are forced to satisfy the conservation of mass, heat, moisture, and momentum through a variational technique. The purpose of this technical report is to provide an overview of the constrained variational analysis method, the architecture of the objective analysis system, along with in-depth information on running the variational analysis codes.

2. Overview of the Constrained Variational Analysis Method

The constrained variational analysis method was developed by Zhang and Lin (1997) for deriving large-scale vertical velocity and advective tendencies from sounding measurements. It is used to process atmospheric soundings of winds, temperature, and water vapor mixing ratio over a network of a small number of stations. Given the inevitable uncertainties in the original data, the basic idea in this objective analysis approach is to adjust these atmospheric state variables by the smallest possible amount to conserve column-integrated mass, moisture, static energy, and momentum. Here, we will briefly review the approach. More details can be seen in Zhang and Lin (1997) and Zhang et al. (2001).

2.1. Theoretical Formulation

From Zhang and Lin (1997), the governing equations of the large-scale atmospheric fields are:

$$\frac{\partial \vec{V}}{\partial t} + \vec{V} \cdot \nabla \vec{V} + \omega \frac{\partial \vec{V}}{\partial p} + f\vec{K} \times \vec{V} + \nabla \phi = -\overline{\nabla \cdot (\vec{V}'\vec{V}')} - \frac{\overline{\partial \omega' \vec{V}'}}{\partial p} \tag{1}$$

$$\frac{\partial s}{\partial t} + \vec{V} \cdot \nabla s + \omega \frac{\partial s}{\partial p} = Qrad + L(C - E) - \overline{\nabla \cdot (\vec{V}' s')} - \frac{\overline{\partial \omega' s'}}{\partial p} + L \frac{\partial q_l}{\partial t} \tag{2}$$

$$\frac{\partial q}{\partial t} + \vec{V} \cdot \nabla q + \omega \frac{\partial q}{\partial p} = E - C - \overline{\nabla \cdot (\vec{V}' q')} - \frac{\overline{\partial \omega' q'}}{\partial p} - \frac{\partial q_l}{\partial t} \tag{3}$$

$$\frac{\partial \omega}{\partial p} + \nabla \cdot \vec{V} = 0 \tag{4}$$

with boundary conditions

$$\omega\Big|_{p=p_s} = \frac{\partial p_s}{\partial t} + \vec{Vs} \cdot \nabla p_s \tag{5}$$

and

$$\omega\Big|_{p=p_0} = 0 \tag{6}$$

where $\vec{V}$ is the wind, $s = C_pT + gz$ is the dry static energy, q is the mixing ratio of water vapor, and p_s is the surface pressure. Large scale is also referred to as grid scale, defined as the size of a sounding array comprising several stations. Prime denotes unresolvable motions of the observational network. Here, Q_{rad} is the net radiative heating rate, C is the condensation of water vapor to rainwater, E is the evaporation of rainwater, and q_l is the cloud liquid water content. Phase changes associated with ice are neglected for simplicity.

Vertical integration of the above equations yields the following constraints:

$$<\nabla \cdot \vec{V}> = -\frac{1}{g}\frac{dp_s}{dt} \tag{7}$$

$$\frac{\partial <q>}{\partial t} + <\nabla \cdot \vec{V}\mathrm{q}> = E_s - P_{rec} - \frac{\partial <q_l>}{\partial t} \tag{8}$$

$$\frac{\partial <s>}{\partial t} + <\nabla \cdot \vec{V}s> = R_{TOA} - R_{SRF} + LP_{rec} + SH + L\frac{\partial <q_l>}{\partial t} \tag{9}$$

$$\frac{\partial <\vec{V}>}{\partial t} + <\nabla \cdot \vec{V}\vec{V}> + f\vec{K} \times <\vec{V}> + \nabla <\phi> = \vec{\tau}_s \tag{10}$$

where

$$< X >= \frac{1}{g}\int_{p_t}^{p_s}(X)dp$$

where R is the net downward radiative flux at Top of the Atmosphere (TOA) and at the surface (SRF), τ_s is the surface wind stress, P_{rec} is precipitation, L is the latent heat of vaporization, SH is the sensible heat flux, E_s is the surface evaporation, and p_t is the TOA pressure.

In the constrained variational analysis method, the atmospheric variables ($\vec{V}$, s, q) are forced to satisfy (7) – (10) with minimum adjustments to direct sounding measurements. The final analysis product is derived by minimizing the cost function:

$$I(t) = \iiint_{p,x,y}\left[\alpha_u(u^* - u_0)^2 + \alpha_v(v^* - v_0)^2 + \alpha_s(s^* - s_0)^2 + \alpha_q(q^* - q_0)^2\right]dxdydp, \qquad (11)$$

with (7) – (10) as strong constraints, where superscript "*" denotes the analyzed data and subscript "0" denotes direct measurements and α is the weighting function related with error estimates in the initial analysis.

2.2. Numerical Implementation

For N stations in the sounding network, each with K layers, we use x_{ik} to denote a state variable at station i and layer k, and use column vector X to denote variable (u, v, s, q) at all grids,

$$X^T = (x_{11}, x_{12}, \ldots, x_{1k}, x_{21}, \ldots, x_{ik}, \ldots, x_{NK}) \qquad (12)$$

where superscript T means transpose. The cost function of (11) can be written as

$$I(t) = (u^* - u_0)^T Q_u (u^* - u_0) + (v^* - v_0)^T Q_v (v^* - v_0) + (s^* - s_0)^T Q_s (s^* - s_0) + (q^* - q_0)^T Q_q (q^* - q_0) \qquad (13)$$

where Q is the weighting matrix related with error covariance of a variable. The analyzed data are subject to the strong constraints of (7) – (10). They can be written in the discrete forms

$$A_{Ps}(\vec{V}_{ik}^*) = (\frac{1}{g}\frac{\partial p_s}{\partial t})_m + <(\nabla \cdot \vec{V}^*)_m>_k = 0 \qquad (14)$$

$$A_q(\vec{V}_{ik}^*, q_{ik}^*) = <(\frac{\partial q^*}{\partial t})_m>_k + <(\nabla \cdot \vec{V}^* q^*)_m>_k - E_s + P_{rec} + <(\frac{\partial q_l}{\partial t})_m>_k = 0 \qquad (15)$$

$$A_s(\overrightarrow{V_{ik}}^*, s_{ik}^*) = <(\frac{\partial s^*}{\partial t})_m>_k + <(\nabla \cdot \vec{V}^* s^*)_m>_k - R_{TOA} + R_{SRF} - LP_{rec} - SH - L<(\frac{\partial q_l}{\partial t})_m>_k = 0 \quad (16)$$

$$A_v(\overrightarrow{V_{ik}}^*, \phi_{ik}^*) = <(\frac{\partial \vec{V}^*}{\partial t})_m>_k + <(\nabla \cdot \vec{V}^* \vec{V}^*)_m>_k + f\vec{K} \times <(\vec{V}^*)_m>_k + <(\nabla \phi^*)_m>_k - \tau_s = 0 \quad (17)$$

where

$$<X> = \sum_{k=1}^{k=K} (X_k) \Delta P_k$$

and subscript m represents average over the area covered by the I stations. Geopotential height can be derived from the virtual temperature analysis using the hydrostatic balance

$$\frac{\partial \phi^*}{\partial p} = -\frac{RT_v^*}{P} \quad (18)$$

The variational equations (Euler-Lagrange equations) for the analyzed variables of u_{ik}^*, v_{ik}^*, q_{ik}^*, s_{ik}^* are :

$$\frac{\partial I(t)}{\partial x_{ik}^*} + \lambda_1(t)\frac{\partial A_{p_s}}{\partial x_{ik}^*} + \lambda_2(t)\frac{\partial A_q}{\partial x_{ik}^*} + \lambda_3(t)\frac{\partial A_s}{\partial x_{ik}^*} + \lambda_4(t)\frac{\partial A_u}{\partial x_{ik}^*} + \lambda_5(t)\frac{\partial A_v}{\partial x_{ik}^*} = 0 \quad (19)$$

where x_{ik}^* stands for any one of u_{ik}^*, v_{ik}^*, q_{ik}^*, s_{ik}^*. Here λ_i is the Lagrange multiplier. Each variable has N × K grids. With a total of four variables and five Lagrange multipliers, the total number of variables to calculate at any given time is 4 × N × K + 5. They are solved from the 4 × N × K equations in (19) and five equations in (14) - (17).

We assume measurement errors at different locations and for different variables uncorrelated. The covariance matrix is then diagonal. The diagonal elements are the reciprocal of error variances $\sigma_{x_{ik}}^2$. Thus, (19) becomes

$$2\sigma_{x_{ik}}^{-2}(x_{ik}^* - x_{0,ik}) + \lambda_1(t)\frac{\partial A_{p_s}}{\partial x_{ik}^*} + \lambda_2(t)\frac{\partial A_q}{\partial x_{ik}^*} + \lambda_3(t)\frac{\partial A_s}{\partial x_{ik}^*} + \lambda_4(t)\frac{\partial A_u}{\partial x_{ik}^*} + \lambda_5(t)\frac{\partial A_v}{\partial x_{ik}^*} = 0 \quad (20)$$

or

$$x_{ik}^{*} = x_{0,ik} - \frac{\sigma_{x_{jk}}^2}{2}[\lambda_1(t)\frac{\partial A_{p_s}}{\partial x_{ik}^{*}} + \lambda_2(t)\frac{\partial A_q}{\partial x_{ik}^{*}} + \lambda_3(t)\frac{\partial A_s}{\partial x_{ik}^{*}} + \lambda_4(t)\frac{\partial A_u}{\partial x_{ik}^{*}} + \lambda_5(t)\frac{\partial A_v}{\partial x_{ik}^{*}}] \quad (21)$$

Numerical calculation of (21) and (14) - (17) is carried out in an iterative mode. The iteration, when described to a single time level, contains three steps. The first step is that the previous estimate or original measurements are used to calculate each partial derivative to x_{ik} on the right-hand side of (21) using the formulas in (14) – (17). A general form of constraint in (14) - (17) can be written as

$$A_Y(x_{ik}^{*}, y_{ik}^{*}) = 0 \quad (22)$$

In an iteration, it can be written as

$$A_Y(x_{ik}^{(m)}, y_{ik}^{(m-1)}) = 0 \quad (23)$$

where m denotes the iteration index. When (21) is written as

$$x_{ik}^{(m)} = x_{0,ik} - \frac{\sigma_{x_{jk}}^2}{2}\sum_{i=1}^{5}\lambda_i(t)B_{i,xk}^{(m-1)} \quad (24)$$

where $B_{i,xk}^{(n-1)}$ are the partial derivatives from the first step, substitution of (24) into (23) yield a linearized set of equations for λ_i. More specifically, the substitution gives

$$A_{Yi}(x_{0,ik} - \frac{\sigma_{x_{jk}}^2}{2}\sum_{i=1}^{5}\lambda_i(t)B_{i,xk}^{(m-1)}, Y_{ik}^{(m-1)}) = 0 \quad (25)$$

Because of the linearity of the above operator, it can be further written as:

$$A_Y(x_{0,ik}, Y_{ik}^{(m-1)}) - \sum_{i=1}^{5}\frac{\sigma_{x_{jk}}^2}{2}\lambda_i(t)A_Y(B_{i,x_{ik}}^{(m-1)}, Y_{ik}^{(m-1)}) = 0 \quad (26)$$

This set of five equations for the five constraints (A_{ps}, A_q, A_s, A_u, A_v) is used to solve for λ_i at any given time. This constitutes the second step in the iteration.

In the third step, the adjustments are calculated by using the newly obtained λ_i in (24). After that, the next iteration is performed.

Because the constraints (14) - (17) contain time derivatives, the actual iteration is carried out simultaneously for all time levels in the field experiment.

3. The Architecture of the ARM Operational Objective Analysis System

This sector provides the details of the implementation of the constrained variational analysis method. Some of them are also described in Zhang et al. (2001). It should be noted that the analysis system is currently designed to process the ARM Single-Column Model (SCM) IOP data at the Southern Great Plains (SGP) site. Thus, some of the implementation methods described below need to be modified according to the actual available data streams and data quality when applied to other locations and/or time periods.

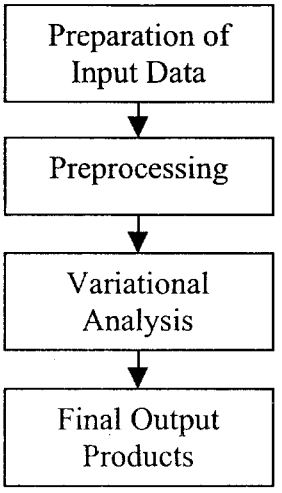

Figure1. Illustration of the structure of the analysis system

Fig. 1 shows the structure of the objective analysis system. It consists of four steps to finish data analysis: (1) preparation of the required raw input data, (2) preprocessing, (3) variational analysis, and (4) output final products. In step (1), all the required data are collected and reorganized from raw datasets to output in a standard format for further analysis. Step (2) includes major quality control of the raw data, averaging the data to form large-scale quantities, filling in missing measurements, and interpolation to consistent observation times. In step (3), the large-scale variables (u, v, T, q) are adjusted by the constrained variational analysis method and the large-scale advective tendencies and vertical velocity are calculated. Step (4) outputs the variables that will be used to force and evaluate SCMs and Cloud Resolving Models (CRMs).

3.1. Preparation of the Required Raw Input Data

3.1.1. Required Input Data

The required input data include the large-scale state variables (u, v, T, q), surface measurements such as surface precipitation rates, surface sensible and latent heat fluxes, surface pressure, surface winds and temperature, surface broadband net radiative flux, net radiation at the TOA, and column total cloud liquid water. Currently, this analysis scheme only adjusts the large-scale state variables. The surface measurements and other data are used as the constraints for the variational analysis.

The large-scale state variables are from the ARM balloon-borne sounding and National Oceanic and Atmospheric Administration (NOAA) wind profiler measurements. There are five sounding stations at the SGP site: the central facility (C1) and four boundary facilities (B1, B4, B5, and B6) (Fig. 2a). During the ARM SCM IOPs, sounding balloons are launched every three-hours to measure the vertical profiles of temperature, water vapor mixing ratio, and winds. There are also 17 NOAA wind profiler stations near the SGP site (crosses in Fig. 2c) taking hourly winds. It is seen from Fig. 2b that all five sounding stations actually overlap with the wind profilers. Around the SGP domain, there is a dense surface measurement network at the ARM SGP site, along with satellite measurements from GOES (Figs. 3a and 3b). These platforms include: (1) the Surface Meteorological Observation Stations (SMOS) that measure surface precipitation, surface pressure, surface winds, temperature and relative humidity; (2) The Energy Budget Bowen Ratio (EBBR) stations that measure surface latent and sensible heat fluxes and surface broadband net radiative flux; (3) The Eddy Correlation Flux Measurement System (ECOR) that provides *in situ* half-hour averages of the surface vertical fluxes of momentum, sensible heat flux, and latent heat flux. (4) The Oklahoma and Kansas mesonet stations (OKM and KAM) that measure surface precipitation, pressure, winds and temperature; (5) The microwave radiometer (MWR) stations that measure the column precipitable water and total cloud liquid water. Radar rainfall data are also available around the SGP domain during certain SCM IOPs. The satellite measurements of clouds and broadband radiative fluxes are available from the 0.5 x 0.5 degree analysis of the Geostationary Operational Environment Satellite (GOES) data (Minnis et al., 1995). In addition to these input data, the analysis system also requires operational analysis of the large-scale state variables from the NOAA mesoscale model RUC (Rapid Update Cycle) as the first guess. The RUC model grids are shown in Fig.

2d. If some other data such as retrievals are available, they can be also easily included into the analysis system.

3.1.2. Files for Preparing the Raw Input Data

Table 1 lists the files that are used to process the raw data in netCDF format from the ARM archive and output the required data in a standard ASCII format for further analysis. These codes are written in FORTRAN77. All missing data are set to –9999.0. Data from different instruments will keep their own observation times at this stage. The output files shown in the third column will be used as the input data for the next step analysis (preprocessing). The variables shown in the third column are described in Table 2.

Table1. File for preparation of the require raw input data

Files	Function	Output
rsonde.for	Process sounding data	**sonde.out**: time, p, lon, lat, u, v, T, dew, RH, alt.
rprofiler.for	Process wind profiler data	**profiler_hour.out**: time, lon, lat, alt, u, v.
rruc.for	Process RUC data	**ruc_model.out**: time, p, Ps, alt, T, RH, u, v.
rebbr.for	Process EBBR data	**surf_ebbr.out**: time, lon, lat, rads, Tsoil, lh, sh.
recor.for	Process ECOR data	**surf_ecor.out**: time, lon, lat, alt, us, vs, ws, Tsair, lh,sh.
rsmos.for	Process SMOS data	**surf_smos.out**: time, lon, lat, alt, us, vs, Tsair, RHs, Ps, prec.
rkam.for	Process KAM data	**surf_kam.out**: time, lon, lat, prec, RHs, Tsair, us, vs.
rokm.for	Process OKM data	**surf_okm.out**: time, lon, lat, prec, Ps, RHs, Tsair, us, vs.
rmwr.for	Process MWR data	**surf_mwr.out**: time, lon, lat, vap, liq.
rgoes.for	Process GOES data	**sat_goes.out**: time, lon, lat, lwt, swt, lwt_clr, ins, swt_clr, cld_low, cld_mid, cld_hgh, cld_tot, cld_top, cld_thick.
rsiros.for	Process SIROS data	**surf_siros.out**: time, lon, lat, lwsup, swsup, lwsdn, swsdn, rads.

Table 2. Available observed variables. Variables necessary for the analysis are marked with "* ".

Variables	Description	Instruments
time (*)	Julian day	All
lon (*)	longitudes	All
lat (*)	latitudes	All
u (m/s) (*)	x-component of horizontal wind	Sounding and wind profilers
v (m/s) (*)	y-component of horizontal wind	Sounding and wind profilers
T (c) (*)	temperature profile	Sounding
RH (%) (*)	relative humidity profile	Sounding
p (mb) (*)	pressure levels	Sounding
dew (c) (*)	dewpoint temperature	Sounding
alt (m) (*)	Altitude	Sounding, wind profilers, and SMOS
rads(w/m2) (*)	net surface radiation	EBBR and SIROS
Tsoil (c)	soil temperature	EBBR
lh (w/m2) (*)	surface latent heat flux	EBBR, ECOR
sh (w/m2) (*)	surface sensible heat flux	EBBR, ECOR
us,vs (m/s) (*)	surface horizontal winds	Sounding, SMOS, KAM, OKM, and ECOR
ws (m/s) (*)	surface vertical wind	ECOR
Tsair (c)	surface air temperature	Sounding, SMOS, KAM, OKM, and ECOR
RHs (%) (*)	surface relative humidity	Sounding, SMOS, KAM, and OKM
Ps (mb) (*)	surface pressure	Sounding, SMOS, and OKM
prec(mm/hr)(*)	surface precipitation rate	SMOS, KAM, OKM, and ABRFC
vap (cm)	column-integrated water vapor	MWR
liq (cm) (*)	column-integrated liquid water	MWR
lwt (w/m2) (*)	TOA net longwave radiative flux	GOES

Table 2. (continue)

Variables	Description	Instruments
swt (w/m2) (*)	TOA net shortwave radiative flux	GOES
lwt_clr (w/m2)	TOA net clear-sky longwave radiative flux	GOES
swt_clr (w/m2)	TOA net clear-sky shortwave radiative flux	GOES
cld_low (%)	low cloud amount	GOES
cld_mid (%)	mid cloud amount	GOES
cld_hgh (%)	high cloud amount	GOES
cld_tot (%)	total cloud amount	GOES
cld_thick (km)	cloud thickness	GOES
cld_top (km)	cloud top	GOES
ins (w/m2)	TOA solar insolation	GOES
lwsup (w/m2)	surface upward longwave radiative flux	SIROS
swsup (w/m2)	surface upward shortwave radiative flux	SIROS
lwsdn (w/m2)	surface downward longwave radiative flux	SIROS
swsdn (w/m2)	surface downward shortwave radiative flux	SIROS

3.2. Preprocessing

Major quality control of the raw data is carried out in this step. Missing data will be filled and data will be averaged to form large-scale quantities. In the end, all upper air state variables (u, v, T, and q) will be interpolated to the final analysis grids and all surface and TOA measurements will be interpolated to the 0.5^0 x 0.5^0 grids within the analysis domain (see Fig. 3b) at consistent analysis times. Results from this step will be used as input for the variational analysis.

3.2.1 Final Analysis Grids

Final analysis grids, as shown in Fig. 2c (dark dots), consist of the five sounding stations, two profiler stations, and six auxiliary grids, which make the analysis domain more symmetric. It is selected to make better use of the hybrid method proposed by Zhang et al. (2001) to calculate the large-scale advective tendencies and vertical velocity. The hybrid method combines the strength of the regular grid method and the line integral method. In the hybrid method, observations are first interpolated onto these analysis grids and then a line-integral method is used in the end to derive the fluxes out of or into the analysis domain. This method has the following advantages: (1) The analysis grids are selected to be as close as possible to the original observational stations, so that an analysis grid is never far away from a measurement station. Since most analysis grid overlap with measurement stations, actual measurements at these grids, if available, will dominate the analysis. (2) The auxiliary grids are added to improve the linear assumption on the line segment at the boundary, making use of measurements that are not located at the domain boundary. (3) This method can also account for the drifting and the time delay of balloons at different stations, since these factors can be included in the interpolation schemes. More details are given in Zhang et al. (2001).

3.2.2 Quality Control

There are four quality control checks used to automatically screen both sounding and wind profiler data.

(1) Maximum and minimum check

The first step is the maximum and minimum check. This procedure will reject those data whose values are larger than their maximum value or smaller than their minimum values. The first two rows in Table 3 show the maximum and minimum values specified for each variables.

(2) Outlier check according to standard deviation

If data departures from the mean larger than four times standard deviation, the data is rejected.

(3) Time variability check

The time variability check is similar to the Grubbs check, which is essentially a high order analysis of variance technique, to check for outliers in the observation, but without being too restrictive so that we retain natural variability in the data. This method is applied with a sliding block of 7 deviation scores. After each test, 6 of them remain and a new deviation score is added to the block. The Grubbs method is used to answer the question: whether this new score contributes an inordinate amount to the variance of the block, so that it ought be considered an outlier. If the answer is no, the score at the front of the block is dropped to make room for the next score to enter the block and be tested. If the answer is yes (an outlier), this outlier score is removed from the block immediately, and the preceding 6 remain unchanged for the next test. The maximum range for time variability is given in the third row in Table 3.

(4) Height-Space check with specified range

This is similar to step (3) instead for height-space check and the last row in Table 3 specifies the maximum range for height-space variability.

It should be noted that, in addition to these procedures, a carefully visual check of all input data is still necessary.

Table 3. Specified ranges for quality check

	time(day)	p(mb)	u(m/s)	v(m/s)	T (c)	Td (c)	RH(%)	H(m)
Max. value	367	1200	70	50	40	40	100	30000
Min. value	0	0	-50	-50	-90	-100	0	0
Time variability	1	3	3	3	3	3	10	100
Height-space variability	0.2	8	3	3	3	3	10	150

3.2.3 Interpolation

After the quality controls, the Cressman scheme (Cressman, 1959) with a four-dimensional length scale of [50 km, 50 km, 50 mb, 6 hours] is used to interpolate the sounding and profiler data onto the final analysis grids. Iteration is carried out for three times. RUC analysis is used

as the background. Since profiler winds are available only in height coordinate, atmospheric temperature and humidity (for the calculation of virtual temperature) are first analyzed using balloon soundings and RUC analysis at the profiler stations. Profiler wind measurements are then converted to pressure levels for subsequent analysis. Considering the temporal and spatial variability of the profiler winds that are less coherent than those from the sounding data, we used a half weight for the profiler data and full weight for the sounding data in the Cressman procedure.

All the constrained terms should be the area-averaged quantities within the analysis domain, since the advective transport terms only describe area-averaged quantities. To avoid biases of using overcrowding measurement stations in some areas, we lay the 0.5^0 x 0.5^0 GOES grids over the analysis domain (Fig. 3b), and then derive the required quantities in each small grid box. If there are actual measurements within the subgrid box, simple arithmetic averaging is used to obtain the subgrid means. Some variables are available from several instruments as shown in Table 2. They are merged in the arithmetic averaging process. In the process, we use a half weight for the data collected from the Oklahoma and Kansas mesonet stations and full weight for the sounding data and SMOS data. If there is no actual measurement in the small box, the Barnes scheme (Barnes, 1964) is used with the length scale of (Lx = 50 km, Ly = 50 km, Lt = 6 hours) to fill the missing data. Domain averages of these quantities are obtained by using values from the 0.5^0 x 0.5^0 grid boxes within the analysis domain. For surface rainfall, we will use radar rainfall data, instead of the data measured from OKM, KAM, and SMOS, whenever it is available. Details about the Cressman and Barnes schemes are described in Cressman (1959) and Barnes (1964), respectively.

3.2.4 Files for Preprocessing

Table 4 lists the required files for preprocessing the raw data from Step 1. The codes are written by IDL. These codes should be run in the order that is shown in Table 4. As mentioned earlier, for those variables that are available from several instruments, they are merged in the arithmetic averaging process.

It should be noted that the file **sub.pro**, which includes all required subroutines, has to be run first before running those files in Table 4.

Table 4. Files for preprocessing. Required input data are shown in parenthesis.

Files	Function	Output
itp_3d_sq	Quality control for sounding data. **(sonde.out)**	**sonde.di5**
plot_sonde_missing	Check missing data in original dataset and the data rejected during the quality control procedure.	
plot_rawsonde_visual	Visual check sounding data after finishing the automatic quality check in **itp_3d_sq.**	
itp_3d_pq	Quality control for wind profiler data. **(profiler_hour.out)**	**Profiler_hour.di5**
itp_3d_pq2	Same as **itp_3d_pq** except that it does 5-point moving smooth for wind profiler data. This is recommended.	**profiler_hour.di5_sm**
plot_prof_missing	Same as **plot_sonde_missing** except for wind profiler data	
plot_rawprof_visual	Same as **plot_rawsonde_visual** except for wind profiler data	
ipt_3d.pro	Allow reading of the sounding, wind profiler, and RUC data **(sonde.out, profiler_hour.out, ruc_model.out)**	
ipt_3d1_loc	Plot locations of the sounding and profiler stations, the RUC model grids, and the final analysis grids. It requires information about the locations.	
itp_3d_ruc	Interpolate RUC data onto the final analysis grids, sounding grids, and wind profiler grids, respectively. **(ruc_model.out)**	**ruc_model.agrid** **ruc_model.prof** **ruc_model.sonde** **ruc_model.agrid_2d** **ruc_model.prof_2d** **ruc_model.sonde_2d**
plot_ruc_visual	Same as **plot_rawsonde_visual** except for the interpolated RUC data	
itp_3d_prof	Convert wind profiler data from height coordinate to pressure coordinate. Temperature and moisture from RUC and sounding at the profiler stations are used to calculate virtual temperature. **(sonde.di5, profiler_hour.di5_sm, ruc_model.prof, ruc_model.sonde, ruc_model.prof_2d, ruc_model.sonde_2d)**	**ruc_prof.prof_tquv**

Table.4. (continue)

Files	Function	Output
plot_prof_visual	Same as **plot_rawsonde_visual** except for the new wind profiler data in **ruc_prof.prof_tquv**	
itp_3d_final	Interpolate sounding and wind profiler data onto the final analysis grids. For wind fields, the sounding and wind profiler data are merged with giving a half weight for the wind profiler data and full weight for the sounding data. **(sonde.di5, ruc_prof.prof_tquv, ruc_model.agrid, ruc_model.sonde, ruc_model.prof, ruc_model.agrid_2d, ruc_model.prof_2d, ruc_model.sonde_2d)**	**analysis.agrid_2d analysis.agird**
plot_final_visual	Same as **plot_rawsonde_visual** except for the data on the final analysis grid.	
ipt_ht	Allow reading of all surface data collected from different instruments **(surf_*.out and sonde.di5)**.	
itp_ht_time	Set time coordinates for different IOPs.	
itp_ht	Interpolate surface data onto the analysis time levels.	
itp_ht_mask	Reject suspicious data through visual check and output required data for further analysis	**surf_*.outn**
ipt_grid_2d	Allow reading of the GOES satellite data. **(sat_goes.out)**	
itp_goes	Interpolate the GOES satellite data onto the analysis time levels	**sat_goes.outn**
ipt_2d	Allow reading of the single layer data processed by itp_ht and itp_goes. **(surf_*.outn and sat_goes.outn)**	
ipt_2d-mix	Merge the header of the single layer data	
itp_2d	Merge the single layer data and interpolate the data onto the 0.5x0.5 grids within the analysis domain	**grid_g.* grid_f.* grid_x.***

3.3 Variational Analysis

As discussed in sector 2.2, numerical calculation of (21) and (14) - (17) is carried out in an iterative mode. The number of iterations is set to 4 in current version. The weights in the cost function are determined by the uncertainties in the instruments and measurements. Zhang and Lin (1997) show that the uncertainties are 0.5 m/s for winds, 0.2 K for temperature, and 3% of the specific humidity based on the estimates by the ARM instrument team. In addition, we also include a component that is related with the observed variance of atmospheric state variables in the sounding data. These standard deviations of the state variables are multiplied by a factor of 20% to account for aliasing errors; they are then added to the instrument and measurement estimates.

Table 5 lists the required files for the constrained variational analysis. These files are written in IDL and they should be run in the order that is shown in Table 5.

Table 5. The files used for the variational analysis. Input data are shown in parenthesis.

Files	Function	Output
3d_put	Allow reading of the multi-layer data and using low pass filtering method to filter the data in time, in pressure, and in space. **(analysis.agrid, sonde.di5, and grid_f.*)**	
2d_put	Allow reading of the single layer data at the 0.5 x 0.5 grids and averaging the data within the analysis domain. **(grid_g.* and grid_x.*)**	
budget_put	Set the budget constrains	
assim	The main program to control the variational analysis	
calc_budget_layer	Calculate budget terms	
assimopt	Output the variational analysis results, including the atmospheric state variables and colimn budget for both before and after the analysis.	***.state** ***.budget_layer** ***.budget_colum**
proc_output	Output the final analysis products	See Table 6.

3.4 Final Output Products

The final outputs from the variational analysis for the single-level time series and multi-layer data are listed below.

Single-level time series:

1. surface precipitation - Prec(mm/hour)
2. surface latent heat flux - LH_(upward_W/m2)
3. surface sensible heat flux - SH_(upward_W/m2)
4. domain averaged surface pressure -Area_Mean_Ps(mb)
5. surface pressure at the ARM central facilty - Central_Facility_Ps(mb)
6. surface air temperature - Ts_Air(C)
7. ground temperature - Tg_Soil(C)
8. surface air relative humidity - Sfc_Air_RH(%)
9. averaged surface wind speed - Srf_wind_speed(m/s)
10. surface u wind - u_wind_(m/s)
11. surface v wind - v_wind(m/s)
12. net downward radiation at the surface - Srf_Net_Dn_Rad(W/m2)
13. net upward radiation at the TOA - TOA_LW_Up(W/m2)
14. net downward shortwave radiative at TOA - TOA_SW_Dn(W/m2)
15. TOA insolation - TOA_Ins(W/m2)
16. low cloud amountfrom GOES- GOES_Lowcld(%)
17. middle cloud amount from GOES -GOES_Midcld(%)
18. high cloud amount from GOES - GOES_Hghcld(%)
19. total cloud amount from GOES- GOES_Totcld(%)
20. cloud thickness from GOES - Cld_Thickness(km)
21. cloud top height from GOES - Cld_Top_ht(km)
22. total cloud liquid water path from microwave radiometers - MWR_Cld_liquid(cm)
23. time change rate of precipitable water - d(Column_H2O)/dt_(mm/hour)
24. column integrated horizontal transport of water vapor - Column_H2O_Advection_(mm/hour)
25. surface evaporation expressed in precipitation unit - Srf_Evaporation_(mm/hour)
26. time change of the column dry static energy - d(Column_Dry_Static_Energy)/dt_(W/m2)
27. column integrated horizontal advection of dry static energy - Column_Dry_Static_Energy_Advection_(W/m2)
28. column integrated radiative heating -Column_Radiative_Heating_(W/m2)
29. column integrated net latent heating - Column_Latent_heating_(W/m2)
30. pressure velocity at the surface- omega_surface_(mb/hr)
31. water vapor mixing ratio at the surface - qs_surface_(kg/kg)
32. dry static energy at the surface - s_surface_(K)
33. precipitable water from microwave radiometers - MWR_precip_water_(cm)
34. surface upward longwawve radiation - Siros_Srf_LWUP_(W/m2)

35. surface downdward longwave radiation - Siros_Srf_LWDN_(W/m2)
36. surface upward shortwave radiation - Siros_Srf_SWUP_(W/m2)
37. surface downward shortwave radiation -Siros_Srf_SWDN_(W/m2)

Multi-Layer Fields of pressure-time cross sections

1. air temperature - Temp_(K)
2. water vapor mixing ratio - H2O_Mixing_Ratio_(g/kg)
3. u wind - u_wind_(m/s)
4. v wind - v_wind_(m/s)
5. pressure vertical velocity - omega_(mb/hour)
6. horizontal wind divergence - Wind_Div_(1/s)
7. horizontal temperature advective tendency - Horizontal_Temp__Advec_(K/hour)
8. vertical temperature advective tendency - Vertical_T_Advec(K/hour)
9. horizontal advective tendency of water vapor - Horizontal_q_Advec_(g/kg/hour)
10. vertical advective tendency of water vapor - Vertical_q_Advec(g/kg/hour)
11. dry static energy - s(Dry_Static_Energy)(K)
12. horizontal advective tendency of dry static energy - Horizontal_s_Advec_(K/hour)
13. vertical advective tendency of dry static energy - Vertical_s_Advec(K/hour)
14. time change rate of dry static energy - ds/dt(K/hour)
15. time change rate of temperature - DT/dt(K/hour)
16. time change rate of water vapor mixing ratio - dq/dt_(g/kg/hour)
17. apparent heating - Q1_(k/hour)
18. apparent water vapor sink -Q2_(K/hour)
19. cloud frequency (cloud amount) from Micro Pulse Lidar- cloud(%).

Acknowledgments. This research was supported primarily under the U. S. Department of Energy Atmospheric Radiation Measurement (ARM) Program. Work at SUNY Stony Brook was supported by ARM grant DE-FG02-98ER62570, and was also supported by NSF under grant ATM9701950. Work at LLNL was performed under the auspices of the U. S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

4. References

Barnes, S. L., 1964: A technique for maximizing details in numerical map analysis. *J. Appl. Meteor.*, **3,** 396-409.

Cressman, G. P., 1959: An operational objective analysis scheme. *Mon. Wea. Rev.*, **87**, 367-374.

Lin, X., and R. H. Johnson, 1994: Heat and moisture budgets and circulation characteristics of a frontal squall line. *J. Atmos. Sci.*, **51**, 1661-1681.

Minnis, P., W. L. Smith, D. P. Garber, J. K. Ayers, and D. R. Doeling, 1995: Cloud properties derived from GOES-7 for spring 1994 ARM Intensive Observing Period using version 1.0.0 of ARM satellite data analysis program. NASA Ref. Publ. 1366, 59pp. [Available from NASA Langley Research Center, Technical Library, MS 185, Hampton, VA 23655-5225.]

O'Brien, J. J., 1970: Alternative solutions to the classical vertical velocity problem. *J. Appl. Meteor.*, **9**, 197-203.

Zhang, M. H., and J. L. Lin, 1997: Constrained variational analysis of sounding data bases on column-integrated budgets of mass, heat, moisture, and momentum: Approach and application to ARM measurements. *J. Atmos. Sci.,* **54**, 1503-1524.

Zhang, M. H., J. L. Lin, R. T. Cederwall, J. J. Yio, and S. C. Xie, 2001: Objective analysis of ARM IOP Data: Method and sensitivity. *Mon. Weather Rev., 129,* 295-311.

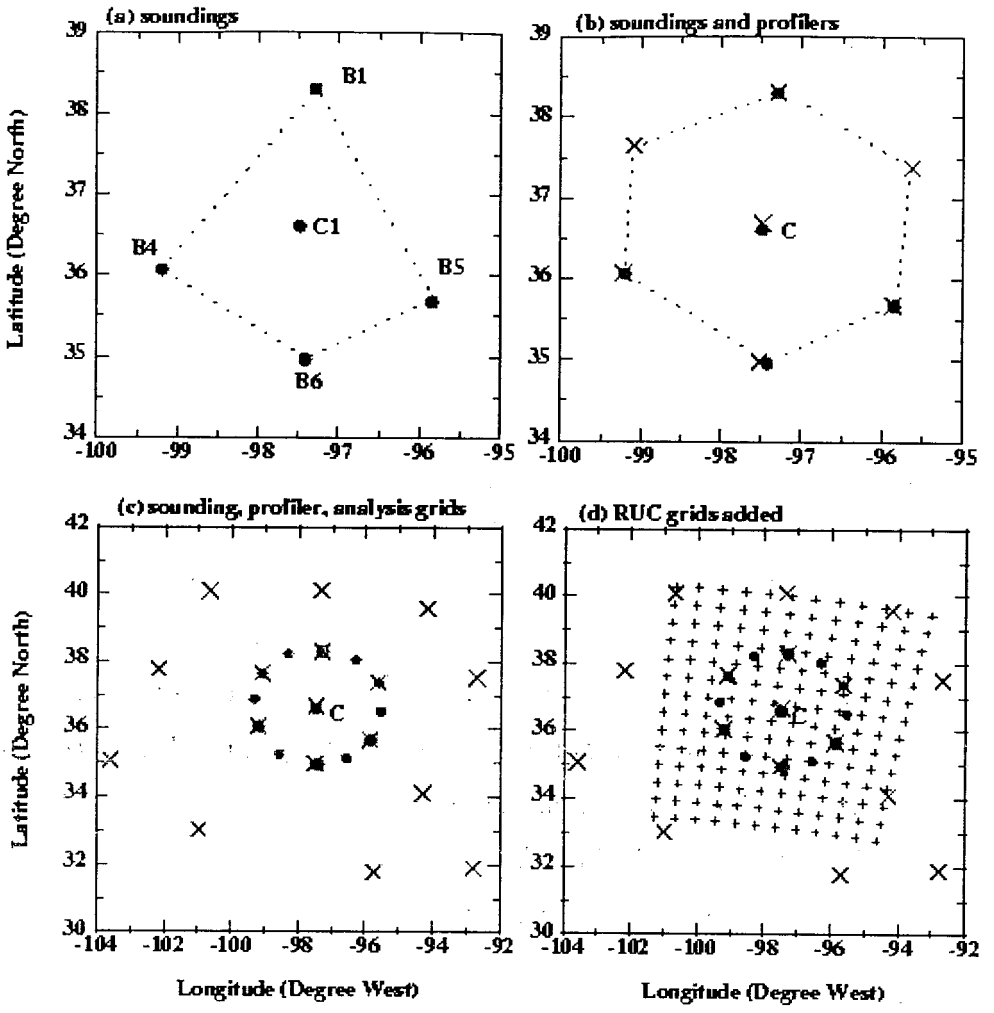

Fig. 2 Locations of the ARM upper-air data streams and the analysis grid points: (a) sounding stations, (b) seven profiler stations (crosses), and (c) the 12 analysis grid points (heavy dots) in the hybrid approach. Also plotted are the nearby profiler stations (crosses). (d) RUC grids overlaid on other grids. (Adapted from Zhang et al., 2001)

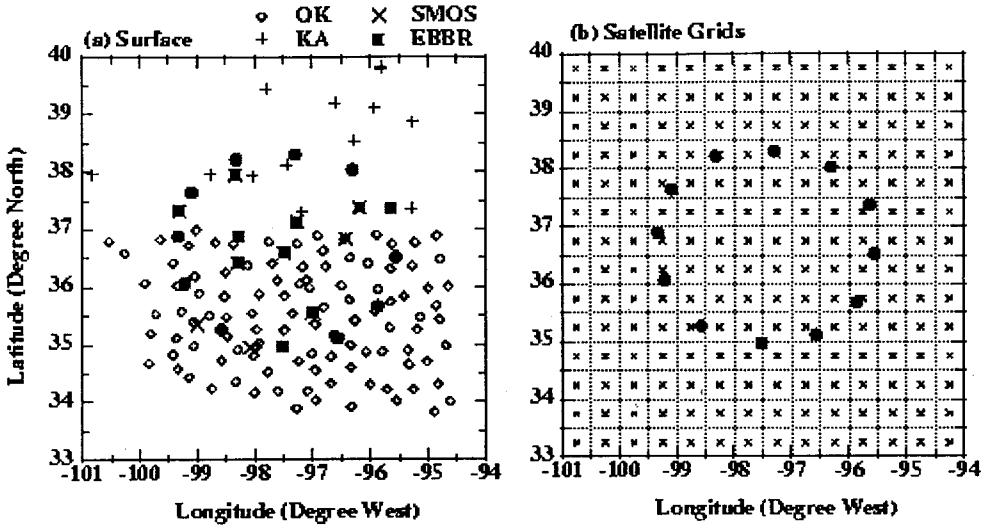

Fig. 3 (a) ARM surface data streams. See text for complete instrument names. (b) GOES grids over the analysis domain. (Adapted from Zhang et al., 2001)